

1	CGACAGTCTTTAGTAGGGAAAGGAGACAAGTGCTAGCTACTGCCGCCCAAGTGAAGGTG	60
61	GGTGAAATTGCTCACTCTTCACCCCACTGACGCTTTTGCGCACCTGGAAAAGCGGTTCCA	120
121	GTTTGCGCCCGTCGCCGCCTTACAGCCGACAGGAGACCAGCGCTACCCAAGTCACGTGGG	180
181	TTCAGCCTGCAGCTTTCTTGCCCCGAAAGGGAATTATCTATAGAGTAAGTATGCTAATCT	240
1		M L I L 4
241	TGACTAAGACTGCAGGAGTTTTTTTTTAAACCATCAAAAAGGAAAGTTTATGAATTTTTAA	300
5	T K T A G V F F K P S K R K V Y E F L R	24
301	GAAGTTTTAATTTTCATCCTGGAACACTATTTCTTCATAAAATAGTATTGGGAATTGAAA	360
25	S F N F H P G T L F L H K I V L G I E T	44
361	CTAGTTGTGATGATACAGCAGCTGCTGTGGTGGATGAAACTGGAAATGTGTTGGGAGAAG	420
45	S C D <u>D</u> T A A A V V D E T G N V L G E A	64
421	CAATACATTCCCAAACCTGAAGTTCATTTAAAAACAGGTGGGATTGTTCTCCAGCAGCTC	480
65	I H S Q T E V H L K T G G I V P P A A Q	84
481	AACAGCTTCACAGAGAAAATATTCAACGAATAGTACAAGAAGCTCTTTCTGCCAGTGGAG	540
85	Q L H R E N I Q R I V Q <u>E</u> A L S A S G V	104
541	TCTCTCCAAGTGACCTCTCAGCAATTGCAACTACCATAAAACCAGGACTTGCTTTAAGCC	600
105	S P S D L S A I A T T I K P G L A L S L	124
601	TGGGAGTGGGCTTATCATTAGCTTACAGCTGGTAGGACAGTTAAAAAAGCCATTCATTC	660
125	G V G L S F S L Q L V G Q L K K P F I P	144
661	CCATTTCATCATATGGAGGCTCATGCACTTACTATTAGGTTGACCAATAAAGTAGAATTTTC	720
145	I <u>H-S-H-M-E-D-A-S-H</u> A L T I R L T N K V E F P	164
721	CTTTTTTAGTTCTTTTGATTTCTGGAGGTCAGTGTCTGTTGGCATTAGTTCAAGGAGTTT	780
165	F L V L L I S G G H C L L A L V Q G V S	184
781	CAGATTTTCTGCTTCTTGAAAAGTCTTTGGACATAGCACCAGGTGACATGCTTGACAAGG	840
185	D F L L L G K S L D I A P G D M L D K V	204
841	TGGCAAGAAGACTTTCTTTAATAAAAACATCCAGAGTGCTCCACCATGAGTGGTGGGAAAG	900
205	A R R L S L I K H P E C S T M S G G K A	224

FIG. 1A

## D0073 CNT

901	CCATAGAACATTTGGCCAAACAAGGAAATAGATTTTCATTTTGACATCAAACCTCCCTTGC	960
225	I E H L A K Q G N R F H F D I K P P L H	244
961	ATCATGCTAAAAATTGTGATTTTCTTTTACTGGACTTCAACACGTTACTGATAAAATAA	1020
245	H A K N C D F S F T G L Q H V T D K I I	264
1021	TAATGAAAAAGGAAAAAGAGGAAGGTATTGAGAAGGGGCAAATCCTGTCTTCAGCAGCAG	1080
265	M K K E K E E G I E K G Q I L S S A A D	284
1081	ACATTGCTGCCACAGTACAGCACACAATGGCATGTCATCTTGTGAAAAGAACACATCGGG	1140
285	I A A T V Q H T M A C H L V K R T H R A	304
1141	CTATTCTGTTTTGTAAGCAGAGAGACTTGTTTACCTCAAAATAATGCAGTACTGGTTGCAT	1200
305	I L F C K Q R D L L P Q N N A V L V A S	324
1201	CTGGTGGTGTGCGCAAGTAACCTTCTATATCCGCAGAGCTCTGGAAATTTTAACAAACGCAA	1260
325	G G V A S N F Y I R R A L E I L T N A T	344
1261	CACAGTGCACCTTTGTTGTGTCCTCCTCCCAGACTATGCACTGATAATGGCATTATGATTG	1320
345	Q C T L L C P P P R L C T D N G I M I A	364
1321	CATGGAATGGTATTGAAAGACTACGTGCTGGCTTGGGCATTTTACATGACATAGAAGGCA	1380
365	W N G I E R L R A G L G I L H D I E G I	384
1381	TCCGCTATGAACCAAAATGTCCTCTTGGAGTAGACATATCAAAAGAAGTTGGAGAAGCTT	1440
385	R Y E P K C P L G V D I S K E V G E A S	404
1441	CCATAAAAGTACCACAATTAAAAATGGAGATATGATTTCTGCTGTTCAAAAAAGTCCCTA	1500
405	I K V P Q L K M E I *	415
1501	AAGGGTCTCACTCTCTGACCTCAGCTGGAGTACAGTAGCCAGATCACAACTCACTGCAAC	1560
1561	CCTGACTTCCTGAACTCAAGAAATCCTCCTGCCTTAGCCTCTTGAATAGCCGGGACTACA	1620
1621	GGTGTGCATGTCCATGCCCAGCCAACTTTATTTCTATTTTTTGTAGAGACAGGCTCTTGC	1680
1681	CATGTTGCCCCGGGCTGGTCCTGAACTGCTGAATTCAAGTGATCCTCCCACCTTGGCCTCC	1740
1741	AGAAGTGCTGGGATTATGGGTGTGAGCCACCATGCCTAGCCAAAATGTTTCTTAAGGTAT	1800
1801	ACATTTTGGGTCTTAGAAGACTTATACATTTGTAATATTTATTACTAAATATCTCAAAGT	1860
1861	ATTACAATAAATGTTACCATGTGAGCTACTTTGAATCAGGCTTCTTGCACACCAATTTAA	1920

FIG. 1B

1921 AAATGTTAACTCTTGATATATACACTAGTTATACCACTCATGTCAGTCAATAAATTTTAA 1980  
1981 GGTTTAAGTGCAGGCCTTTGTTTACAGAAATCCTAATTTTTTGAAACCATAACTCTGACC 2040  
2041 TGACACTAAATTCCTGTAGACATGCTAAGGAAAATCTGCTTAGTATCGAGATCAAGAACT 2100  
2101 TCCATTCAAAAAGATTATTTCAGTTATGTTATTTGCATATTACCATTGTTAAAAATAAAAA 2160  
2161 AATTTTTTAAAAGATGAAAAAAAAAAAAAAAAAAAAAAAAA 2197

**FIG. 1C**

		1	50
MP-1	(1)	-----MLIL	
gi 2583127	(1)	MVRLFLTLSIPAISRFNLYPGISILARNNNSLRLQKHKLKTKTPTFSLIS	
gi 7495111	(1)	-----	
gi 4980638	(1)	-----	
GCP_HELPY	(1)	-----	
		51	100
MP-1	(5)	TKTAGVFFKPSKRKVYEFLLRSFNHFG-TLFLHKIVLGIETSCDDTAAAV	↓
gi 2583127	(51)	PSSSPNFQTRFYSTETRISLSPYSENPNFDDNLVVLGIETSCDDTAAAV	
gi 7495111	(1)	-----MNIPKILNNNLVLKRIFCRNYSVKVLGIETSCDDTAVAI	
gi 4980638	(1)	-----MRVLGIETSCDETAAVAV	
GCP_HELPY	(1)	-----MILSTESSCDDSSIAL	
		101	150
MP-1	(54)	V-DETNVLGEAIIHSQTEVHLKTTGGIVPPAAQQLHRENIQRIVQEALSAS	↓
gi 2583127	(101)	V-SPFNHLSS---SCRAELLVQYGGVAPKQAEAAHSRVIDKVVQDALDKA	
gi 7495111	(40)	VNEKREILSSE-RYTERAIQRQGGINPSVCALQHRENLPRLIEKCLNDA	
gi 4980638	(18)	L-DDGKNVVVNFTVSQIEVHQKFGGVVPEVAARHLLKNLPILLKKAFEKV	
GCP_HELPY	(17)	TRIEDAQLIAHFKISQEKHHSSYGGVPELASRLHAENLPLLLERIKISL	
		151	200
MP-1	(103)	GVSPSDLSAIIATTIKPGILALSIGVGLSFSLQLVGQLKKPFIPIHHMEAAHA	
gi 2583127	(147)	NLTEKDLSAVAVTIGPGLSLCLRVGVRKARRVAGNFSLPIVGVMHMEAAHA	
gi 7495111	(89)	GTSPKDLDAVAVITVTPGLVIALKEGISAAIGFAKKHRLPLIPVHHMRAHA	
gi 4980638	(67)	PP--ETVDVVAATYGPGLIGALLVGLSAAKGLAISLEKPFVGVNHVEAAHV	
GCP_HELPY	(67)	NKDFSKIKAIATINQPLSVTLIEGLMMAKALSLSLNLPLILEDHLRGHV	
		201	250
MP-1	(153)	LTIRLTN-KVEFPFLVLLISGGHCLLALVQGVSDFLLLGSLSLDIAPGDML	
gi 2583127	(197)	LVARLVEQELSFPFMALLISGGHNLLVLAHKLQYQTQLGTTVDDAIGCAF	
gi 7495111	(139)	LSILLVDDSVRFPPSAVLLISGGHALISVAEDVEKFKLYGQSVSGSPGECI	
gi 4980638	(115)	QAVFLANPDLKPPPLVVLVMSGGHTQLMKVDEDYSMEVLGETLDDSGAEAF	
GCP_HELPY	(117)	YSLFINEKQTCMPLSVLLVSGGHSLLILEARDYENIKIVATSLDDSGESF	
		251	300
MP-1	(202)	DKVARRLSLIKHPCESTMSGGAIEHLAKQGNR--FHFDIKPPLHHAKNK	
gi 2583127	(247)	DKTAKWLGDMHRS-----GPAVEELALEGDA--KSVKFNVPKMYHKDC	
gi 7495111	(189)	DKVARQLGDLGSEFDG-IHVCAAVEILASRASAD-GHLRYPILFPNVPKA	
gi 4980638	(165)	DKVARLLGLGYP-----GCPVIDRVAKKGDP--EKYSFPRPMLDDDSY	
GCP_HELPY	(167)	DKVSKMLDLGYP-----GCPIVEKLALDYRHPNEPLMFPIPLKNSPNL	
		301	350
MP-1	(250)	DFSFTGLQHVTDKIIMKKEKEEGIEKGQILSSAADIAATVOHTMACHLVK	
gi 2583127	(290)	NFSYAGLKTQVRLAIEAKE-----IRNRADIAASFQAVLHLEE	
gi 7495111	(237)	NMNFQDIKGSYLNLLERLRKNSE-----TSIDIPDFCASLONTVARHISS	
gi 4980638	(206)	NFSFAGLKTSLVLYFLQREK-----GYKVEDVAASFQKAVVDILVE	
GCP_HELPY	(210)	AFSFSGLKNAVRLEVEKNAPNLN-----EAIKQKIGYHFSAAIEHLIQ	
		351	400
MP-1	(300)	RTHRAILFCKQRDLLPQNNAVLVASGGVASNFYIRRALEILTNAQTCTLL	
gi 2583127	(330)	KCERAIDWALELEPSIKH---MVISGGVASNKYVRLRLNNIVENKNLKL	
gi 7495111	(282)	KLHIFFESLSEQEKLPKQ---LVIGGGVAANQYIFGAISKLSAAHNVTTI	
gi 4980638	(246)	KTFR-----LARNLGIRK---IAFVGGVAANSMLREEVRKRAERWNYEVF	
GCP_HELPY	(254)	QTKR-----YFKIKRPKIFGIVGGASQNLALRKAFENLCDAFDCKLV	

FIG. 2A

		401		450
MP-1	(350)	CPPPRLCTDNGIMIAWNGIERLR-AGLGILHDIEGIRYEPKCPLGVDISK		
gi 2583127	(377)	CPPPSLCTDNGVMVAWTGLEHFR-VGRYDPPPPATEPEDYVYDLRPRWPL		
gi 7495111	(329)	KVLLSLCTDNAEMIAYSGLLMLVNRSEAIWWRPNIPDTIYAHARSDIGT		
gi 4980638	(288)	FPPLELCTDNALMVAKAGYEKAK-RGMFSPLSLNADPNLNV-----		
GCP_HELPY	(296)	LAPLEFCSDNAAMIGRSSLEAYQ-KKRFVPLEKANISPRTLKSF-----		
		451		493
MP-1	(399)	EVGEASIKVPQLKMEI-----		
gi 2583127	(426)	GEEYAKGRSEARSMRTARIHPSLTSIIRADSLQQQTQT-----		
gi 7495111	(379)	DASSEIIDTPRRKLVSTIHGTERIRFRNLDDFKKPKSPKTTE		
gi 4980638	(328)	-----		
GCP_HELPY	(341)	-----		

FIG. 2B

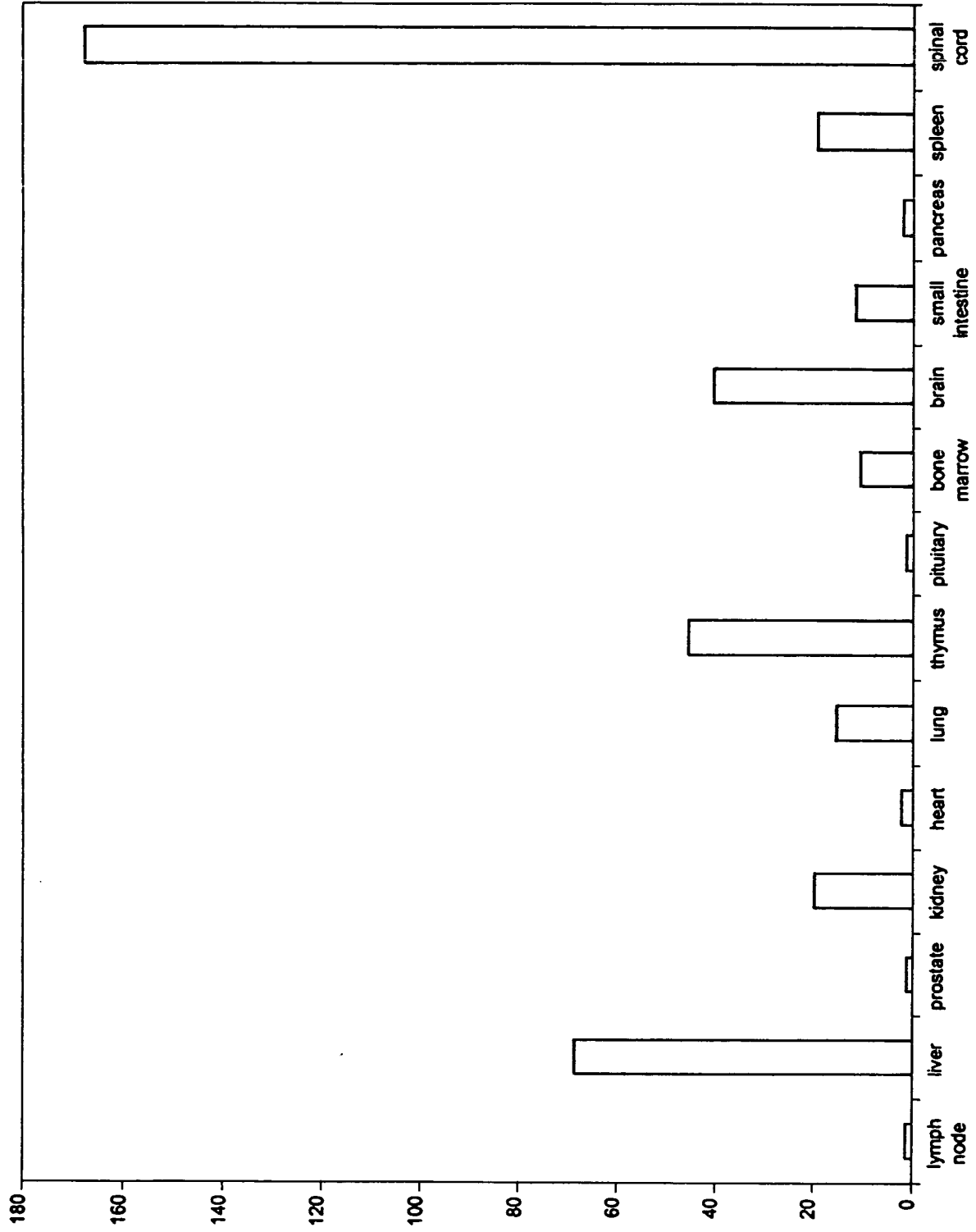


FIG. 3

		1		*50
MP-1	(1)	MLILTKTAGVFFKPSKRKVYEFLRSFNHFGTLFLHKIVLGIETSCDDTA		
HYPD	(1)	-----		
		51		*100
MP-1	(51)	AAVVDETGNVLGEAIIHSQTEVHLKTGGIVPPAAQQLHRENIQRIVQEALS		
HYPD	(1)	-----		
		101		* 150
MP-1	(101)	ASGVSPSDLSAIATTIKPGLALSIGVGLSFSLQLVGQLKKPFIPHHMEA		
HYPD	(1)	-----		
		151		200
MP-1	(151)	HALTIRLTNKVEFPFLVLLISGGHCLLA <del>LV</del> <del>GV</del> <del>SDF</del> <del>LL</del> LGKSLDIAPGDM		
HYPD	(1)	-----MRIL <del>V</del> <del>L</del> <del>G</del> <del>V</del> <del>G</del> <del>N</del> <del>I</del> <del>L</del> TDEAIGVRIVEA		
		201		250
MP-1	(201)	<del>LD</del> KVARRLSLIKHP <del>EC</del> S <del>TS</del> MSGGKA <del>IE</del> H <del>LA</del> KQCN <del>RF</del> <del>FD</del> IKPPLHHA <del>KN</del> CD		
HYPD	(26)	<del>LE</del> QRYILPDYVEILDGG <del>TA</del> G---M <del>EL</del> LGDMA <del>NR</del> <del>DE</del> L <del>IT</del> ADAIVSK <del>KN</del> AP		
		251		300
MP-1	(251)	FSFTG <del>IL</del> QHVTDKIIMKKEKEEG <del>IE</del> K <del>CO</del> ILSSAADIAATVQHTMACH <del>LV</del> KR		
HYPD	(72)	GTMMI <del>LR</del> DEEVPALFTN---K <del>IS</del> PH <del>OL</del> GLADVLSA <del>LR</del> FTGEFPKK <del>IT</del> LV		
		301		350
MP-1	(301)	THRAILFCKQRD <del>IL</del> <del>LP</del> ONNAVLVASGGVASNFYIRRA <del>LE</del> ILT <del>NA</del> TQCTLLC		
HYPD	(118)	GVI <del>P</del> ESLEPHIG <del>LT</del> <del>ET</del> VEAMIEPALEQVLAALRESG <del>VE</del> AI <del>PR</del> SDS-----		
		351		400
MP-1	(351)	PPRLCTDNGIMIAWNGIERLRAGLGILHDIEGIRYEPKCPLGVDISKEV		
HYPD	(163)	-----		
		401	414	
MP-1	(401)	GEASIKVPQLKMEI		
HYPD	(163)	-----		

FIG. 4

		1	50
MP-1	(1)	MLILTKTAGVFFKPSKRKVYEF LRSFNHPC	TLFLHKIVLGIETSCDDTA
11641265	(1)	MLILTKTAGVFFKPSKRKVYEF LRSFNHPE	TLFLHKIVLGIETSCDDTA
		51	100
MP-1	(51)	AAVVDETGNVLGEAIHSQTEVHLKTGGIVPPAAQQLHRENIQRIVQEALS	
11641265	(51)	AAVVDETGNVLGEAIHSQTEVHLKTGGIVPPAAQQLHRENIQRIVQEALS	
		101	150
MP-1	(101)	ASGVSPSDLSAIATTIKPGLALSLGVGLSFSLQLVGQLKKPFIPIHHMEA	
11641265	(101)	ASGVSPSDLSAIATTIKPGLALSLGVGLSFSLQLVGQLKKPFIPIHHMEA	
		151	200
MP-1	(151)	HALTIRLTNKVEFPFLVLLISGGHCLLALVQGVSDFLLL GKSLDIAPGDM	
11641265	(151)	HALTIRLTNKVEFPFLVLLISGGHCLLALVQGVSDFLLL GKSLDIAPGDM	
		201	250
MP-1	(201)	LDKVARRLSLIKHPECSTMSGGKAIEHLAKQGNRFHFDIKPPLHHAKNCD	
11641265	(201)	LDKVARRLSLIKHPECSTMSGGKAIEHLAKQGNRFHFDIKPPLHHAKNCD	
		251	300
MP-1	(251)	FSFTGLQHVTDKIIMKKEKEEGI	-----EK
11641265	(251)	FSFTGLQHVTDKIIMKKEKEEGI	FLISKVEQINIPGLCLKIAAHFCRYEK
		301	350
MP-1	(276)	GQILSSAADIAATVQHTMACHLVKRTHRAILFCKQRDLLPQNNAVLVASG	
11641265	(301)	GQILSSAADIAATVQHTMACHLVKRTHRAILFCKQRDLLPQNNAVLVASG	
		351	400
MP-1	(326)	GVASNFYIRRALEILT NATQCTLLCPPPRLCTDNGIMIAWNGIERLRAGL	
11641265	(351)	GVASNFYIRRALEILT NATQCTLLCPPPRLCTDNGIMIAWNGIERLRGGL	
		401	439
MP-1	(376)	GILHDIEGIRYEPKCPLGVDISKEVGEASIKVPQLKMEI	
11641265	(401)	GILHDIEGIRYEPKCPLGVDISKEVGEASIKVPQLKMEI	

FIG. 5



mp1	1	MLILTKTAGVFFKPSKRKVYEFLLRSFNHPGTLFLHKIVLGIETSCDDT MLILTKTAGVFFKPSKRKVYEFLLRSFNHPGTLFLHKIVLGIETSCDDT MLILTKTAGVFFKPSKRKVYEFLLRSFNHPGTLFLHKIVLGIETSCDDT
AC013468_6	-9850	acataaaaggggtactaaagtgttaatatccgactccaagtgagaatgga ttttcacogtttaccagataattgggtatacgcctttaatttgtaacggaac gacgtgtaattttaaagattataaattttttaaatttaagatatatttta
mp1	50	AAAVVDETG NVLGEAIHSQTEVHL AAAVVDETG NVLGEAIHSQTEVHL AAAVVDETG NVLGEAIHSQTEVHL
AC013468_6	-9703	gggggggagagtgaggactcaggctAAGTAAGTA Intron 1 cccttaacgattgactacacatat <2-----[9629 : 3771] attgggtatatggaaaatcatatta
mp1	74	TGGIVPPAAQQLHRENIQRIVQEALSASGVSPSDLSAIATTIKPG TGGIVPPAAQQLHRENIQRIVQEALSASGVSPSDLSAIATTIKPG TGGIVPPAAQQLHRENIQRIVQEALSASGVSPSDLSAIATTIKPG
AC013468_6	-3773	CAGAaggagccggccccagaaaccagcggtgaggtcagctgagaaaacg -2> cggttccccaatagaaatagttaactcgggtccgatcctccctacg atgtttaatagtcgaattaaaaaatttctactatccaatatcaaaa
mp1	120	LALSLGVGLSFSLQLVGQLKKPFIPHHMEAHALTIRLTNKVEFPFLVL LALSLGVGLSFSLQLVGQLKKPFIPHHMEAHALTIRLTNKVEFPFLVL LALSLGVGLSFSLQLVGQLKKPFIPHHMEAHALTIRLTNKVEFPFLVL
AC013468_6	-3634	cgtaagggttttatccggctaactacaccaggcgcaaaataaaggctcttgc tctgtgtgtctgtattgataacttctaatacactctgtcaatatctttt ttacgagcaatcaggaagaagactcctttgggtattttggctaaatttatt
mp1	169	LISGGHCLLALVQGVSDFLLL GKSLDIAPGDM LDK LISGGHCLLALVQGVSDFLLL GKSLDIAPGDM LDK LISGGHCLLALVQGVSDFLLL GKSLDIAPGDM LDK
AC013468_6	-3487	tatggctctgtgcgggtgtcccgattgagcggaacga ttcggagttcttagtcattttgactatccgattaa gttatctggaataatattgttagtgcaaatcgctcg
mp1	204	VARRLSLIKHP ECSTMSGGKAIEHLA VARRLSLIKHP ECSTMSGGKAIEHLA VARRLSLIKHP ECSTMSGGKAIEHLA
AC013468_6	-3382	GTAATTA Intron 2 TAGggaacttaaccgttaaaggagagctg <0-----[3382 : 2556]-0>tcgggtcttaacagcctgggactaatc gaaattaaatagcccgttgacaatgc
mp1	230	KQGNRFHFDIKPPLHHAKNCDFSFTGLQHVTDKIIMKKEKEE KQGNRFHFDIKPPLHHAKNCDFSFTGLQHVTDKIIMKKEKEE KQGNRFHFDIKPPLHHAKNCDFSFTGLQHVTDKIIMKKEKEE
AC013468_6	-2477	acgaatctgaacctccgaatgttttagccccgagaaaaaagagg aagagtataacctaacaagatctcgtaatcaattttaaaaaa aaatattttccatcgttttttttttatactttaagagaaga

FIG. 6A

mp1	272		IEKGQILSSAADIAATVQHTMAC
			IEKGQILSSAADIAATVQHTMAC
		G:G[ggt]	IEKGQILSSAADIAATVQHTMAC
AC013468_6	-2351	GGTATATT Intron 3	CAGGTagagcacttgggaggagccaagt
		<1-----[2350 : 2275]-1>	taagattcccatccctaactcg
			tgggacgtaaacttcaagcagat
mp1	296	HLVKRTHRAILEFCKQRDLLPONNAVL	
		HLVKRTHRAILEFCKQRDLLPONNAVL	
		HLVKRTHRAILEFCKQRDLLPONNAVL	
AC013468_6	-2203	ccgaaaccgacttacagttccaaggcGTAAGTT Intron 4	CAG
		attagcagctttgaagattcaaactt<0-----[2125 : 1190]-0>	
		ttgaaatggttggttgacgatattaag	
mp1	322	VASGGVASNFYIRRALEILTNATQCTLLCPPPRLCTDNGIMIA	
		VASGGVASNFYIRRALEILTNATQCTLLCPPPRLCTDNGIMIA	
		VASGGVASNFYIRRALEILTNATQCTLLCPPPRLCTDNGIMIA	
AC013468_6	-1189	ggtggggaattacagcgataagactatttcccactagagaaaag	
		tccggtcgatatggctattcaccagcttgcccgtgcaagtttc	
		tatttcacatccatgataataagctgggtttcaactttctgta	
mp1	365		NGIERLRAGLGILHDIEGIRYEP
			NGIERLRAGLGILHDIEGIRYEP
		W:W[tgg]	NGIERLRAGLGILHDIEGIRYEP
AC013468_6	-1060	TGGTAAGCC Intron 5	TAGGagagaccggtgatcgaggactgc
		<2-----[1058 : 935]-2>	agtagtgcggtgtaatagtgaaac
			tttaaattcgctatcaaccctaa
mp1	389	K	
		K	
		K	
AC013468_6	-864	a	
		a	
		g	

FIG. 6B

11641265	1	MLILTKTAGVFFKPSKRKVEFLRSFNHPETLFLHKIVLGIETSCDDT MLILTKTAGVFFKPSKRKVEFLRSFNHP TLFLHKIVLGIETSCDDT MLILTKTAGVFFKPSKRKVEFLRSFNHPGTLFLHKIVLGIETSCDDT
AC013468_6	-9850	acataaaggggttactaaagtgttaatatccgactccaagtgagaatgga ttttcaccggtttaccagataattggtatacgctttaattgtacggaac gacgtgtaatttaaagattataattttttaaatttaaagatatatttta
11641265	50	AAAVVDETGNVLGEAIHSQTEVHL AAAVVDETGNVLGEAIHSQTEVHL AAAVVDETGNVLGEAIHSQTEVHL
AC013468_6	-9703	gggggggagagtgaggactcaggctAAGTAAGTA Intron 1 cccttaacgattgactacacatat <2-----[9629 · 3771] attggtatatggaaaatcatatta
11641265	74	TGGIVPPAAQQLHRENIQRIVQEALSASGVSPSDLSAIATTIKPG TGGIVPPAAQQLHRENIQRIVQEALSASGVSPSDLSAIATTIKPG TGGIVPPAAQQLHRENIQRIVQEALSASGVSPSDLSAIATTIKPG
AC013468_6	-3773	CAGAaggagccggcccccagaaccagcggctgaggtcagctgagaaaacg -2> cggttccccaatagaaatagttaactccggtccgatccctccctacg atgtttaatagtcatttaaaaaatttctactatccaatatcaaaa
11641265	120	LALSLGVGLSFSLQLVGQLKKPFIPiHHMEAHALTIRLTNKVEFPFLVL LALSLGVGLSFSLQLVGQLKKPFIPiHHMEAHALTIRLTNKVEFPFLVL LALSLGVGLSFSLQLVGQLKKPFIPiHHMEAHALTIRLTNKVEFPFLVL
AC013468_6	-3634	cgtacgggttttatccggctaactacaccaggcgcaaaataaaggtcttgc tctgtgtgtctgtattgataacttctaatacactctgtcaatatctttt ttacgagcaatcaggaagaagactctttggtattttggctaaatttatt
11641265	169	LISGGHCLLALVQGVSDFLLLGKSILDIAPGDMLDK LISGGHCLLALVQGVSDFLLLGKSILDIAPGDMLDK LISGGHCLLALVQGVSDFLLLGKSILDIAPGDMLDK
AC013468_6	-3487	tatggctctgtgcggtgtcccgattgagcggacga ttcggagttcttagtcattttgactatccgattaa ggtatctggaataatattgttagtgcaaatcgtcg
11641265	204	VARRLSLIKHPECSTMSGGKAIEHLA VARRLSLIKHPECSTMSGGKAIEHLA VARRLSLIKHPECSTMSGGKAIEHLA
AC013468_6	-3382	GTAATTA Intron 2 TAGggaacttaaccgttaaaggagagctg <0-----[3382 : 2556]-0>tcggtcttaacagcctgggactaatc gaaattaaatagcccgttgacaatgc
11641265	230	KQGNRFHFDIKPPLHHAKNCDFSTGLQHVTDKIIMKKEKEEGIFLISK KQGNRFHFDIKPPLHHAKNCDFSTGLQHVTDKIIMKKEKEEGIFLISK KQGNRFHFDIKPPLHHAKNCDFSTGLQHVTDKIIMKKEKEEGIFLISK
AC013468_6	-2477	acgaatctgaacctccgaatgttttagccccgagaaaaaagagggatcaaaa aagagtataatacctaacaagatctcgtaataatttaaaaaagttttga aaatatttccatogttttatttttttataactttaagagaagatatatta

FRAME SHIFT HERE

FIG. 7A

11641265	279	VEQINIPGLCLKIAAHFCRYEKGQILSSAADIAATVQHTMACHLVKRTH VEQINIPGLCLKIAAHFCRY KGQILSSAADIAATVQHTMACHLVKRTH VEQINIPGLCLKIAAHFCRY!KGQILSSAADIAATVQHTMACHLVKRTH	
AC013468_6	-2330	ggcaaacggttcaaggcttat4agcacttgggaggagccaagtccgaaac taatatcgtgtatccatgga agattcccatcccctaactcgattagca tagatttagcaaatttcogt ggacgtaaacttcaagcagatttgaaat	
11641265	328	RAILFCKQORDLLPQNAVL RAILFCKQORDLLPQNAVL RAILFCKQORDLLPQNAVL	VASGGVA VASGGVA VASGGVA
AC013468_6	-2182	cgacttacagttccaaggcGTAAGTT Intron 3 gctttgaagattcaaactt<0-----[2125 : 1190]-0>tcoggtc gttgttgacgatattaag	CAGggtgggg tatttca
11641265	354	SNFYIRRALEILTNATQCTLLCPPPRLCTDNGIMIA SNFYIRRALEILTNATQCTLLCPPPRLCTDNGIMIA SNFYIRRALEILTNATQCTLLCPPPRLCTDNGIMIA	
AC013468_6	-1168	aattacagcgataagactatttcccactagagaaag gatatggctattcaccagottgccogtgcaagtttc tcctccatgataataagctgggtttcaactttctgta	
11641265	390		NGIERLRGGLGILHDIEGIRYEP NGIERLR GLGILHDIEGIRYEP NGIERLRAGLGILHDIEGIRYEP
AC013468_6	-1060	TGGTAAGCC Intron 4 TAGGagagaccggtgatogaggactgc <2-----[1058 : 935]-2> agtagtgogtgtaatagtgaac tttaaattcgctatcaaccctaa	
11641265	414	K K K	
AC013468_6	-864	a a g	

FIG. 7B

<b><u>Protein</u></b>	<b><u>Genbank ID</u></b>	<b><u>Identities</u></b>	<b><u>Similarities</u></b>
Arabidopsis O-sialoglycoprotein endopeptidase protein	gi 2583127	30.8%	36%
Caenorhabditis glycoproteinase family member protein	gi 7495111	26.6%	41%
Thermotoga secreted metalloendopeptidase Gcp protein	gi 4980638	27.2%	39%
Helicobacter O-sialoglycoprotein endopeptidase GCP_HELPY protein	gi 2499846	22.5%	34%

**FIG. 8**

D0073 CNT



FIG. 9

## D0073 CNT

		1	50
MP-1	(1)	CGACAGTCTTTAGTAGGGAAAGGAGACAAGTGCTAGCTACTGCCGCCAA	
gi 11641264	(1)	-----	
		51	100
MP-1	(51)	GTGGAAGGTGGGTGAAATTGCTCACTCTTCACCCCACTGACGCTTTTGCG	
gi 11641264	(1)	-----	
		101	150
MP-1	(101)	CACCTGGAAAAGCGGTTCCAGTTTGCGCCCGTCGCCGCTTACAGCCGAC	
gi 11641264	(1)	-----	
		151	200
MP-1	(151)	AGGAGACCAGCGCTACCCAAGTCACGTGGGTTTCAGCTGCAGCTTTCTTG	
gi 11641264	(1)	-----	
		201	250
MP-1	(201)	GCCCGAAAGGGAATTATCTATAGAGTAAGTATGCTAATCTTGACTAAGAC	
gi 11641264	(1)	-----CAGGAATTATCTATAGAGTAAGTATGCTAATCTTGACTAAGAC	
		251	300
MP-1	(251)	TGCAGGAGTTTTTTTAAACCATCAAAAAGGAAAGTTTATGAATTTTAA	
gi 11641264	(44)	TGCAGGAGTTTTTTTAAACCATCAAAAAGGAAAGTTTATGAATTTTAA	
		301	350
MP-1	(301)	GAAGTTTAAATTTTCATCCTGGAACACTATTTCTTCATAAAATAGTATTG	
gi 11641264	(94)	GAAGTTTAAATTTTCATCCTGGAACACTATTTCTTCATAAAATAGTATTG	
		351	400
MP-1	(351)	GGAATTGAAACTAGTTGTGATGATACAGCAGCTGCTGTGGTGGATGAAAC	
gi 11641264	(144)	GGAATTGAAACTAGTTGTGATGATACAGCAGCTGCTGTGGTGGATGAAAC	
		401	450
MP-1	(401)	TGGAAATGTGTGGGAGAAGCAATACATTCCCAAAGTGAAGTTCATTAA	
gi 11641264	(194)	TGGAAATGTGTGGGAGAAGCAATACATTCCCAAAGTGAAGTTCATTAA	
		451	500
MP-1	(451)	AAACAGGTGGGATTGTTCTCCAGCAGCTCAACAGCTTCACAGAGAAAAT	
gi 11641264	(244)	AAACAGGTGGGATTGTTCTCCAGCAGCTCAACAGCTTCACAGAGAAAAT	
		501	550
MP-1	(501)	ATTCAACGAATAGTACAAGAAGCTCTTTCTGCCAGTGGAGTCTCTCCAAG	
gi 11641264	(294)	ATTCAACGAATAGTACAAGAAGCTCTTTCTGCCAGTGGAGTCTCTCCAAG	
		551	600
MP-1	(551)	TGACCTCTCAGCAATTGCAACTACCATAAAACCAGGACTTGCTTTAAGCC	
gi 11641264	(344)	TGACCTCTCAGCAATTGCAACTACCATAAAACCAGGACTTGCTTTAAGCC	
		601	650
MP-1	(601)	TGGGAGTGGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAAAG	
gi 11641264	(394)	TGGGAGTGGGCTTATCATTTAGCTTACAGCTGGTAGGACAGTTAAAAAG	
		651	700
MP-1	(651)	CCATTCATTCCCATTTCATCATATGGAGGCTCATGCACTTACTATTAGGTT	
gi 11641264	(444)	CCATTCATTCCCATTTCATCATATGGAGGCTCATGCACTTACTATTAGGTT	
		701	750
MP-1	(701)	GACCAATAAAGTAGAATTTCCTTTTTAGTTCTTTTGATTCTGGAGGTC	
gi 11641264	(494)	GACCAATAAAGTAGAATTTCCTTTTTAGTTCTTTTGATTCTGGAGGTC	
		751	800
MP-1	(751)	ACTGCTCTGTTGGCATTAGTTCAAGGAGTTTCAGATTTCTGCTTCTTGGA	
gi 11641264	(544)	ACTGCTCTGTTGGCATTAGTTCAAGGAGTTTCAGATTTCTGCTTCTTGGA	
		801	850
MP-1	(801)	AAGTCTTTGGACATAGCACCAGGTGACATGCTTGACAAGGTGGCAAGAAG	
gi 11641264	(594)	AAGTCTTTGGACATAGCACCAGGTGACATGCTTGACAAGGTGGCAAGAAG	

FIG. 10A

## D0073 CNT

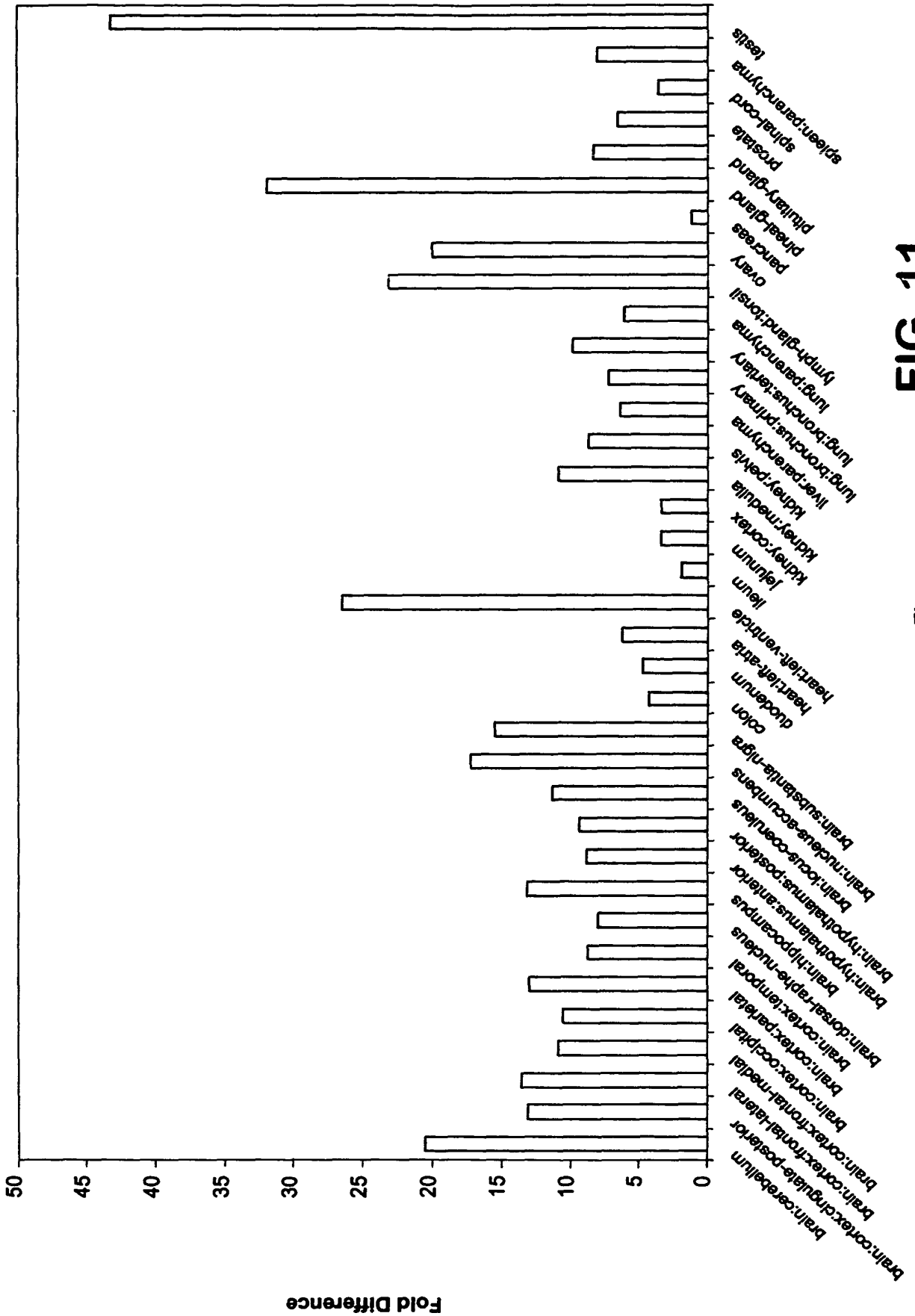
			851	900
MP-1	(851)		ACTTTCTTTAATAAAACATCCAGAGTGCTCCACCATGAGTGGTGGGAAAG	
gi 11641264	(644)		ACTTTCTTTAATAAAACATCCAGAGTGCTCCACCATGAGTGGTGGGAAAG	
			901	950
MP-1	(901)		CCATAGAACATTTGGCCAAACAAGGAAATAGATTTCATTTTGACATCAAA	
gi 11641264	(694)		CCATAGAACATTTGGCCAAACAAGGAAATAGATTTCATTTTGACATCAAA	
			951	1000
MP-1	(951)		CCTCCCTTGCATCATGCTAAAAATTGTGATTTTCTTTTACTGGACTTCA	
gi 11641264	(744)		CCTCCCTTGCATCATGCTAAAAATTGTGATTTTCTTTTACTGGACTTCA	
			1001	1050
MP-1	(1001)		ACACGTTACTGATAAAATAATAATGAAAAAGGAAAAAGAGGAAGGTAT--	
gi 11641264	(794)		ACACGTTACTGATAAAATAATAATGAAAAAGGAAAAAGAGGAAGGTATAT	
			1051	1100
MP-1	(1049)		-----	
gi 11641264	(844)		TTCTAATTAGTAAAGTTGAACAGATAAAATATTCCTGGATTGTGCCTAAAA	
			1101	1150
MP-1	(1050)		-----TGAGAAGGGGCAAATCCTGTCTTCAGC	
gi 11641264	(894)		ATAGCTGCTCATTTCTGCAGGTAATGAGAAGGGGCAAATCCTGTCTTCAGC	
			1151	1200
MP-1	(1076)		AGCAGACATTGCTGCCACAGTACAGCACACAATGGCATGTCATCTTGTGA	
gi 11641264	(944)		AGCAGACATTGCTGCCACAGTACAGCACACAATGGCATGTCATCTTGTGA	
			1201	1250
MP-1	(1126)		AAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGAGACTTGTACCT	
gi 11641264	(994)		AAAGAACACATCGGGCTATTCTGTTTGTAAAGCAGAGAGACTTGTACCT	
			1251	1300
MP-1	(1176)		CAAAATAATGCAGTACTGGTTGCATCTGGTGGTGTGCGCAAGTAACTTCTA	
gi 11641264	(1044)		CAAAATAATGCAGTACTGGTTGCATCTGGTGGTGTGCGCAAGTAACTTCTA	
			1301	1350
MP-1	(1226)		TATCCGCAGAGCTCTGGAAATTTTAACAAACGCAACACAGTGCACCTTTGT	
gi 11641264	(1094)		TATCCGCAGAGCTCTGGAAATTTTAACAAACGCAACACAGTGCACCTTTGT	
			1351	1400
MP-1	(1276)		TGTGTCTCTCTCCAGACTATGCACTGATAATGGCATTATGATTGCATGG	
gi 11641264	(1144)		TGTGTCTCTCTCCAGACTATGCACTGATAATGGCATTATGATTGCATGG	
			1401	1450
MP-1	(1326)		AATGGTATTGAAAGACTACGTGCTGGCTTGGGCATTTTACATGACATAGA	
gi 11641264	(1194)		AATGGTATTGAAAGACTACGTGCTGGCTTGGGCATTTTACATGACATAGA	
			1451	1500
MP-1	(1376)		AGGCATCCGCTATGAACCAAAATGTCTCTTGGAGTAGACATATCAAAAG	
gi 11641264	(1244)		AGGCATCCGCTATGAACCAAAATGTCTCTTGGAGTAGACATATCAAAAG	
			1501	1550
MP-1	(1426)		AAGTTGGAGAAGCTTCCATAAAAGTACCACAATTAAAAATGGAGATATGA	
gi 11641264	(1294)		AAGTTGGAGAAGCTTCCATAAAAGTACCACAATTAAAAATGGAGATATGA	
			1551	1600
MP-1	(1476)		TTTCTGCTGTTCAAAAAAGTCCCTAAAGGCTCTCACTCTCTGACCTCAGC	
gi 11641264	(1344)		TTTCTGCTGTTCAAAAAAGTCCCTAAAGGCTCTCACTCTCTGACCTCAGC	
			1601	1650
MP-1	(1526)		TGGAGTACAGTAGCCAGATCACAACCTCACTGCAACCTTGACTTCCTGAAC	
gi 11641264	(1388)		-----	
			1651	1700
MP-1	(1576)		TCAAGAAATCCTCCTGCCTTAGCCTCTTGAATAGCCGGGACTACAGGTGT	
gi 11641264	(1388)		-----	

FIG. 10B



		1701		1750
MP-1	(1626)	GCATGTCCATGCCAGCCAACTTTATTCTATTTTGTAGAGACAGGCT		
gi 11641264	(1388)	-----		
		1751		1800
MP-1	(1676)	CTTGCCATGTTGCCCGGGCTGGTCCTGAAGTCTGAATTCAGTGATCCT		
gi 11641264	(1388)	-----		
		1801		1850
MP-1	(1726)	CCCACCTTGGCCTCCAGAAGTGCTGGGATTATGGGTGTGAGCCACCATGC		
gi 11641264	(1388)	-----		
		1851		1900
MP-1	(1776)	CTAGCCAAAATGTTTCTTAAGGTATACATTTGGGTCTTAGAAGACTTAT		
gi 11641264	(1388)	-----		
		1901		1950
MP-1	(1826)	ACATTGTGAATATTTATTACTAAATATCTCAAAGTATTACAATAAATGTT		
gi 11641264	(1388)	-----		
		1951		2000
MP-1	(1876)	ACCATGTGAGCTACTTTGAATCAGGCTTCTGCACACCAATTTAAAAATG		
gi 11641264	(1388)	-----		
		2001		2050
MP-1	(1926)	TTAACTCTTGATATATACACTAGTTATACCACTCATGTCAGTCAATAAAT		
gi 11641264	(1388)	-----		
		2051		2100
MP-1	(1976)	TTTAAGGTTTAAGTGCAGGCCTTTGTTTACAGAAATCCTAATTTTGTAA		
gi 11641264	(1388)	-----		
		2101		2150
MP-1	(2026)	ACCATAACTCTGACCTGACACTAAATTCCTGTAGACATGCTAAGGAAAT		
gi 11641264	(1388)	-----		
		2151		2200
MP-1	(2076)	CTGCTTAGTATCGAGATCAAGAACTTCCATTCAAAAAGATTATTCAGTTA		
gi 11641264	(1388)	-----		
		2201		2250
MP-1	(2126)	TGTTATTTGCATATTACCATTGTTAAAAATAAAAAATTTTAAAGATG		
gi 11641264	(1388)	-----		
		2251	2272	
MP-1	(2176)	AAAAAAAAAAAAAAAAAAAAA		
gi 11641264	(1388)	-----		

FIG. 10C



**FIG. 11**